B16

**PCT** WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12Q 1/02, 1/68

A1

(11) International Publication Number: WO 96/23899

(43) International Publication Date: 8 August 1996 (08.08.96)

(21) International Application Number: PCT/US95/09589

(22) International Filing Date: 27 July 1995 (27.07.95)

(30) Priority Data:
08/381,800 1 February 1995 (01.02.95) US

(71) Applicant: UNIVERSITY OF MASSACHUSETTS MEDICAL CENTER [US/US]; 55 Lake Avenue North, Worchester, MA 01655 (US).

(72) Inventors: FRITZ, Christian, C.; 195 South Quinsigamond Avenue, Shrewsbury, MA 01545 (US). GREEN, Michael, R.; 368 Green Street, Boylston, MA 01505 (US).

(74) Agents: CLARK, Paul, T. et al.; Fish & Richardson P.C., 225 Franklin Street, Boston, MA 02110 (US).

(81) Designated States: AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TT, UA, UG, UZ, VN, ARIPO patent (KE, MW, SD, SZ, UG), European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

**Published**
*With international search report.*

(54) Title: METHODS OF SELECTING A RANDOM PEPTIDE THAT BINDS TO A TARGET PROTEIN

(57) Abstract

A method of isolating a random peptide-scaffold peptide fusion protein that binds a predetermined target protein is disclosed. First, a library of nucleotide sequences encoding random amino acid sequences are generated and cloned into a first expression vector to make a library test vector. The nucleotide sequence is expressed within a yeast cell as a test fusion protein containing a transcriptional activation domain, a random peptide domain, and a scaffold peptide domain. The DNA sequence encoding a target protein cloned into a second expression vector provides a target vector. The expressed target fusion protein contains a DNA-binding domain and a target protein. Test fusion protein-target fusion protein interaction is detected according to a two-hybrid protein-protein interaction assay. Also disclosed are an HIV-1 Rev protein binding peptide isolated by the method of the invention and methods of using the Rev interacting protein.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | |
|---|---|
| AM | Armenia |
| AT | Austria |
| AU | Australia |
| BB | Barbados |
| BE | Belgium |
| BF | Burkina Faso |
| BG | Bulgaria |
| BJ | Benin |
| BR | Brazil |
| BY | Belarus |
| CA | Canada |
| CF | Central African Republic |
| CG | Congo |
| CH | Switzerland |
| CI | Côte d'Ivoire |
| CM | Cameroon |
| CN | China |
| CS | Czechoslovakia |
| CZ | Czech Republic |
| DE | Germany |
| DK | Denmark |
| EE | Estonia |
| ES | Spain |
| FI | Finland |
| FR | France |
| GA | Gabon |
| GB | United Kingdom |
| GE | Georgia |
| GN | Guinea |
| GR | Greece |
| HU | Hungary |
| IE | Ireland |
| IT | Italy |
| JP | Japan |
| KE | Kenya |
| KG | Kyrgystan |
| KP | Democratic People's Republic of Korea |
| KR | Republic of Korea |
| KZ | Kazakhstan |
| LI | Liechtenstein |
| LK | Sri Lanka |
| LR | Liberia |
| LT | Lithuania |
| LU | Luxembourg |
| LV | Latvia |
| MC | Monaco |
| MD | Republic of Moldova |
| MG | Madagascar |
| ML | Mali |
| MN | Mongolia |
| MR | Mauritania |
| MW | Malawi |
| MX | Mexico |
| NE | Niger |
| NL | Netherlands |
| NO | Norway |
| NZ | New Zealand |
| PL | Poland |
| PT | Portugal |
| RO | Romania |
| RU | Russian Federation |
| SD | Sudan |
| SE | Sweden |
| SG | Singapore |
| SI | Slovenia |
| SK | Slovakia |
| SN | Senegal |
| SZ | Swaziland |
| TD | Chad |
| TG | Togo |
| TJ | Tajikistan |
| TT | Trinidad and Tobago |
| UA | Ukraine |
| UG | Uganda |
| US | United States of America |
| UZ | Uzbekistan |
| VN | Viet Nam |

BNSDOCID: <WO___9623899A1_I_>

# METHODS OF SELECTING A RANDOM PEPTIDE THAT BINDS TO A TARGET PROTEIN

## Statement as to Federally Sponsored Research

This invention was made at least in part with funds from the Federal government, and the government therefore has rights in the invention.

## Background of the Invention

The invention relates to a method of selecting a fusion protein which binds to a target protein of interest within a host cell. The invention also relates to a polypeptide that binds to the human immunodeficiency virus (HIV)-1 Rev protein, and methods of inhibiting HIV-1 production using the polypeptide.

Protein-protein interactions within a cell are the basis for countless physiological control mechanisms. Viral control mechanisms are the targets of searches for the cellular host proteins with which a viral protein interacts as part of its function in the cell.

The "two-hybrid" assay, is a genetic screen originally described by Fields et al. (Fields S. et al. (1989) *Nature*, 340:245-246; Fields, S. and Sternglanz, R. (1994) *Trends in Genetics*, 10:286-292) to test protein-target protein interactions. The "two-hybrid" selection system utilizes the interaction between two proteins of interest to reconstitute an active transcription factor from two fusion proteins; one containing a DNA binding domain fused to the protein of interest, and the other containing a transcriptional activation region fused to a cDNA library (Chien C. T. et al. (1991) *P.N.A.S. USA*, 88:9578-9582; Dalton, S. et al. (1992) *Cell*, 68:597-612). Test protein-target protein interaction is indicated by expression of a marker gene operably linked to a promoter activated by the reconstituted transcriptional factor.

Though useful, the two hybrid system cannot assay all protein-protein interactions. According to Fields, S. and Sternglanz, R. (1994), supra, proteins that cannot fold correctly within the cell are not suitable for selection by this system.

Viruses are obligate intracellular parasites and depend on protein-protein interactions between viral and host cell proteins to carry out functions vital to their life cycle. Blocking or inhibiting these critical interactions is a therapeutic strategy that may have profound effects on the viral pathogen and little effect on the host cell. Furthermore, as mutation of the viral protein may be constrained by the requirement for functional interaction with host factors, such a strategy may be resistent to viral escape mutations.

Current strategies to inhibit replication of HIV are commonly aimed at the virus' reverse transcriptase using nucleoside analogues like AZT, ddC and ddI. However the complex life cycle of this retrovirus offers unique targets of intervention in the form of small regulatory proteins that are now being intensively studied. One of these proteins, Rev, is a regulator of virion protein expression.

Rev was first identified in proviral mutants that were unable to replicate because they did not produce detectable levels of Env and Gag proteins (Sodroski et al. (1986) *Nature*, 321:412-417; Feinberg et al. (1986) *Cell*, 46:807-817). It is now clear that Rev induces the synthesis of these structural proteins by allowing the appearance of their mRNAs in the cytoplasm (Malim, M. et al. (1989) *Nature*, 338:254-257; Emerman, M. et al. (1989) *Cell*, 57:1155-1165). Since the mRNAs coding for Gag, Env, and Pol contain introns with sequences to which the Rev protein specifically binds, these mRNAs are normally translocated to the cytoplasm

for translation to their respective proteins during HIV infection.

In the absence of Rev protein function, the mRNAs encoding Gag, Env, and Pol remain in the nucleus, are not expressed, and HIV-1 replication is inhibited. As no cellular protein is known to perform a similar function, and because the concentration of the Rev protein must exceed a threshold for function (Pomerantz, R.J. et al. (1992) *J. Virology*, 66:1809-1813), Rev is considered a prime target for specific antiviral agents (Cohen, J. (1993) *Science*, 260:1257; Rosen, C.A. (1992) *AIDS Research and Human Retroviruses*, 8:175-181; Duan, L. et al. (1994) *P.N.A.S. USA*, 91:5075-5079).

Rev is a small 116 amino acid phosphoprotein that is located predominantly in the nucleus/nucleolus of human cells infected with HIV-1 (Cullen, B. R. et al. (1988) *J. Virol.*, 62:2498-2501; Felber, B. K. et al. (1989) *P.N.A.S. USA*, 86:1495-1499). Extensive mutagenesis has revealed that Rev contains at least two domains that are necessary for function: 1) a basic RNA binding domain at the N-terminus that binds to a specific RNA sequence (RRE, Rev Response Element) present in all unspliced HIV mRNAs, and 2) a C-terminal effector domain (Malim, M. et al. (1989) *Cell*, 58:205-214; Olsen, H.S. et al. (1990) *Genes and Dev.*, 4:1357-1364) necessary for function. The effector domain was defined by mutations, e.g., M10, that not only abolish Rev function but at the same time create transdominant negative mutants, i.e., mutants able to repress wild type Rev function when added in a molar excess (Malim, M. et al. (1989), *supra*). The M10 mutation does not affect the normal multimerization of Rev nor the ability of Rev to bind RNA, suggesting that the region of Rev containing the M10 mutation interacts with one or more cellular proteins (Malim, M. et al. (1991) *J. Virol.*, 65:4248-

4254) and that the interaction with a putative cellular target is essential for Rev function.

## Summary of the Invention

In general, the invention features a method of selecting a random amino acid sequence that binds to a target protein in a host cell, e.g., a yeast cell. The invention is based on the unexpected discovery that a stably folded polypeptide domain (a "scaffold") fused to the random amino acid sequence enhances such selection. In addition, the invention features a random peptide detected by the method of the invention, which is not naturally expressed in the cell from which its transcript was isolated, and which unexpectedly exhibited specific binding to the Rev protein effector domain.

It is an object of the invention to enhance the ability to isolate random peptides which bind to a target protein of interest by expressing the random peptide within a known stable 3-dimensional structure, or scaffold, as a fusion protein. This fusion protein is expressed from a nucleic acid sequence encoding a random sequence of amino acids covalently linked to a scaffold peptide. The stably folded structure of the scaffold enhances and/or enables binding of the random peptide to the target protein. Binding of the random peptide-scaffold peptide fusion to the target protein of interest is monitored by a two-hybrid selection method (Fields, S. et al. (1989), *supra*).

In general, the scaffold is a peptide sequence that stably folds into a 3-dimensional structure and which, when fused to a random amino acid sequence of the invention, enhances presentation of the random sequence to the target protein. Specific embodiments of the scaffold peptide include nucleotide sequences which encode and allow translation of an Adh C-terminus (SEQ ID

NO:3); the B1 domain of G protein (Gronenborn, A.M. (1991) *Science*, 253:657-661); a zinc finger domain (Lee, M.S. (1989), *Science*, 245:635); the B domain of protein A from *S. aureus* (Gouda et al. (1992) *Biochemistry* 31:9665-9672); the tenth type III module of fibronectin (Baron et al. (1991) *Biochemistry* 31:2068-2073); and cyclophilin (Ke et al. (1991) *PNAS USA* 88:9483-9487). The random peptide can be linked at either end, or preferably in the middle, and most preferably in external loops of the scaffold, making the random peptide accessible to the target protein for binding.

The invention further features a non-natural fusion protein, RIP (Rev interacting protein, SEQ ID NO:2), isolated by expressing a non-naturally translated peptide fused to a scaffold peptide domain, which fusion peptide binds to the HIV-1 Rev protein and inhibits its function.

In addition, the invention features a substantially pure DNA sequence (SEQ ID NO:4), referred to herein as the "R32 nucleotide sequence," and the deduced amino acid sequence (SEQ ID NO:5). This sequence does not include any polylinker or scaffold sequences.

The invention also features a Rev protein binding protein encoded by a DNA sequence of SEQ ID NO:4 linked to a scaffold peptide-encoding sequence.

The invention also features a vector containing, in the order of transcription, a nucleotide sequence encoding one member of a two-hybrid binding peptide pair (i.e. a transcriptional activation domain (Gal4 or *Herpes virus* VP16, for example); a nucleotide sequence encoding a random peptide sequence to be tested; and a nucleotide sequence encoding a scaffold peptide domain.

In addition, the invention features a second vector containing, in the order of transcription, a nucleotide sequence encoding a DNA binding domain (Gal4

or LexA, for example); and a nucleotide sequence encoding a target protein. It is understood by one of ordinary skill in the art that the transcriptional activation domain or the complementary DNA binding domain of a two-hybrid binding peptide pair can be varied in position with its respective fusion protein as necessary to improve transcriptional activation.

One of the vectors of the two-hybrid system contains a nuclear localization signal to allow translocation of the complex of interacting fusion proteins into the nucleus for interaction with the transcriptional activation sequences of the marker gene. It is also noted that vectors used in selecting a random peptide in the context of a scaffold peptide domain can encode the random peptide 5′ to the scaffold peptide domain (in the order of transcription), or the random peptide-encoding sequence can be inserted into the coding sequence of the scaffold peptide domain.

The invention also features an antibody to the RIP protein. The antibody may be a monoclonal antibody or a polyclonal antibody. For use in assays of the invention, the anti-RIP antibody, or the RIP itself, may also be labeled for detection such as with a fluorescent moiety covalently attached to the antibody, or peptide, without interfering with antibody binding to the RIP.

In addition, the invention features an assay for the qualitative or quantitative detection of Rev protein in a cellular lysate, by inducing a population of cells containing or suspected of containing HIV to proliferate, thereby inducing HIV to replicate and produce Rev protein. The cells are lysed and the proteins of the lysate are contacted with RIP, e.g., immobilized on a solid support, to allow binding of Rev. The presence of the RIP-Rev complex is indicated by a decrease in the

detection signal observed when Rev is present compared to a control sample having no Rev protein.

The invention further features a method for screening candidate agents that inhibit HIV replication by inhibiting Rev production. Human cells containing actively replicating HIV are contacted with the candidate agent in an appropriate culture medium. As measured, e.g., by the RIP/anti-RIP antibody assay, a decrease in the cellular concentration of Rev in treated cells compared to a control without the candidate agent, indicates that the agent is potentially useful for inhibiting HIV replication in human cells.

The invention also features a method of inhibiting HIV reproduction in lymphocytes of a mammal (preferably a human) by introducing DNA encoding RIP into the lymphocytes where the DNA is expressed and RIP is produced in amounts sufficient to inhibit HIV reproduction. Moreover, the invention features a method of inhibiting the production of HIV-1 in human cells, e.g., therapeutically or prophylactically, by administering the RIP protein to the cells using gene therapy techniques.

By "random peptide" or "random amino acid sequence," is meant any peptide that does, or does not, naturally exist in a cell. For example, a random peptide according to the invention can be generated by the translation of a nucleotide sequence that does not encode a naturally occurring full-length protein, or by synthetic methods. The nucleotide sequence encoding a random peptide is encoded, for example, by a cDNA sequence isolated by reverse transcription of a cellular sense or antisense mRNA transcript, or by a synthetic oligonucleotide containing amino acid codons linked in random order. Examples of a nucleotide sequence encoding a random peptide include, but are not limited to: 1)

cellular antisense mRNA reverse transcribed into cDNA then expressed as a polypeptide which binds to a target protein of interest, and 2) a polypeptide generated by the expression of a nucleic acid sequence encoding a randomly generated amino acid sequence.

A "stably folded peptide" is a peptide that is resistant to denaturation, e.g., thermal denaturation, in a cell.

By "reduced HIV-1 protein production in cells expressing the virus" is meant a decrease in the expression of, for example, essential proteins Gag, Env, and Pol in RIP-containing cells compared to the normal expression in the same cells lacking RIP.

By "sufficient concentration of HIV-1 Rev binding peptide" is meant a concentration of RIP in a cell above the threshold necessary to inhibit translocation the Gag, Env, and Pol mRNAs from the nucleus into the cytoplasm.

By "substantially pure random peptide-scaffold peptide fusion protein" is meant that the random peptide-scaffold peptide fusion protein provided by the invention is at least 60%, by weight, free from the proteins and naturally-occurring organic molecules with which it is naturally associated. Preferably, the preparation is at least 75%, more preferably at least 90%, by weight, random peptide-scaffold peptide fusion protein. A substantially pure random peptide-scaffold peptide fusion protein may be obtained, for example, by expression of the recombinant nucleic acid of the library test vector encoding it, or by chemically synthesizing the protein. Purity can be measured by any appropriate method, e.g., column chromatography, polyacrylamide gel electrophoresis, or HPLC analysis.

By "substantially identical amino acid sequence" is meant an amino acid sequence that differs from another sequence only by conservative amino acid substitutions,

for example, substitution of one amino acid for another of the same class (e.g., valine for glycine or arginine for lysine), or by one or more non-conservative amino acid substitutions, deletions, or insertions at positions of the amino acid sequence that do not destroy the biological activity of the random peptide-scaffold peptide fusion protein.

By "operably linked" means that selected DNA is in proximity with transcriptional and translational regulatory sequences to allow these sequences to regulate expression of the selected DNA.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the detailed description, and from the claims.

## Brief Description of the Drawings

Fig. 1 is a diagram of the results of HIV antigen production and Rev binding experiments comparing characteristics of wild type Rev protein and various mutant Rev proteins. Data for HIV-1 gtat and p24 antigen production were obtained from Malim, et al. (1991), *supra*. RIP-Rev binding experiments were performed by the method of the invention.

Fig. 2 is a diagram of R32 clones used to evaluate regions necessary for Rev binding activity.

Fig. 3A is a diagram showing various N-terminal deletions of the R32 nucleotide sequence starting at the *Eco*RI site. The *Xho*I site was filled in with α-thionucleotides to achieve unidirectional deletions according to standard methods (Henikoff (1984) *Gene*, 28:351-359).

Fig. 3B is a diagram showing the structure of the initial test vector construct and of active test vector constructs after bidirectional exonuclease III deletions. The deletions were started from the *Xba*I site.

Fig. 4 is a representation of the nucleotide and deduced amino acid sequences of the Rev interacting protein, RIP. The RIP coding sequence consists of the downstream 175 nucleotides of the originally isolated R32 sequence (from the human MCM2 homologue in antisense orientation) linked to DNA linker sequences from the cloning vector cDNA, 15 nucleotides from the yeast ADH gene, and 108 nucleotides encoding 35 amino acids plus a stop codon (SEQ ID NO:3) from the Adh C-terminus. The shaded portions have been found to be dispensable for binding to Rev in C-terminal deletion experiments, but were included for cloning convenience in the experiments expressing RIP in mammalian cells.

Fig. 5 is a graph showing the inhibition of HIV-1 gag p24 antigen production by increasing amounts of RIP-encoding DNA introduced into HeLa cells by cotransfection with HIV-1 infectious molecular clone, pNL4-3 (Adachi, A. et al. (1986) *J. Virol.* 59:284-291).

Figs. 6A and 6B are graphs showing the inhibition of HIV production in CEM lymphocytes by RIP protein (Fig. 6A); and showing that RIP-induced inhibition is specific to HIV-1 Rev since RIP does not affect expression of another protein, alkaline phosphatase, in the CEM

lymphocytes (Fig. 6B). The results also show that inhibition of HIV-1 production by RIP is not caused by non-specific cell toxicity of RIP since the expression of a control gene, alkaline phosphatase, is unaffected.

Fig. 7 is a graph showing the specific and dosage dependent inhibition of HIV-1 production in H9 cells by RIP. The results show that the inhibitory effect of RIP is due to expression of the RIP peptide and is not caused by inhibition of translation due to an antisense effect.

Fig. 8 is a diagram of a method of assaying for the presence of Rev in a cell lysate by contacting the cellular and viral proteins of the cell lysate with RIP immobilized on a solid support. Detection is accomplished by the relative amounts of anti-RIP antibody that is excluded from binding to the support due to the presence of the Rev-RIP complex.

## Detailed Description

The following examples are set forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to select, in a host cell, a fusion protein containing a random amino acid sequence fused to a scaffold protein domain, which fusion protein binds a specific target protein. Also disclosed is such a fusion protein, RIP, and methods of using the same. Such disclosures are not intended to limit the scope of the invention.

As originally presented, the two-hybrid system exploits the ability of a pair of interacting proteins to bring a transcription activation domain into close proximity with a DNA-binding site that regulates the expression of an adjacent reporter rene (Fields, S. and Song, O.-K. (1989) *Nature*, 340:245-246). In general, the two-hybrid assay is performed in yeast cells and requires the construction of hybrid genes to encode 1) a DNA

binding domain fused to a protein X; and 2) an activation domain fused to a protein Y. When both hybrid proteins are present, the noncovalent interaction of X and Y bridges the interaction of the DNA binding domain with the DNA binding site, and the interaction of the activation domain with the transciptional machinery, thereby activating transcription of the reporter gene. In the general assay, to select a protein Y, that binds to a target protein X, Y is encoded by a pool of plasmids in which total cDNA or genomic DNA is ligated to the activation domain (Fields, S. and Sternglanz, R. (1994) *Trends in Genetics*, 10:286-292).

The following examples provide support for the discovery that a scaffold peptide domain fused to a random peptide enhances binding selection of the random peptide in a host cell, e.g., a yeast cell.

## Examples

### Example 1: A Method of Isolating a Random Peptide-Scaffold Fusion Protein That Binds a Specific Target Protein

Selection of a random peptide in the "two-hybrid" assay, is limited by the ability of the random peptides to fold correctly within the cell (Fields, S. and Sternglanz, R. (1994) *Trends in Genetics*, 10:286-292). Most random peptide sequences are unlikely to fold into a stable compact protein. As described in the examples, the unexpected finding that a Rev interacting protein required an Adh scaffold sequence (SEQ ID NO:3) for binding demonstrates a need for a stably folded scaffold peptide to be fused to the random amino acid sequence to be selected.

To provide a useful scaffold peptide domain, the Adh transcription termination sequence was expressed as a read-through sequence which accompanied transcription of

the random amino acid sequence to be selected by the method. The observed binding requirement for the Adh transcription termination sequence indicated that the Adh sequence was necessary as a molecular scaffold that stabilized the entire fusion protein, and thus the random peptide. Thus, the 35 amino acids, LVKSPIKVVGLSTLPEIYEKMEKGQIVRYVVDTSK (SEQ ID NO:3) from the Adh transcription termination sequence are required to be present in the expressed fusion protein to serve as the scaffold for the purpose of structural presentation of a fused random peptide in a two hybrid selection system.

Another scaffold protein sequence useful in the method of the invention is the B1 domain of the Streptococcal protein G. This protein domain is a single immunoglobulin binding domain, and the folded structure has been extensively studied (Gronenborn, A.M. et al. (1991) *Science*, 253:657-661), but never used in the two-hybrid system. The B1 protein domain forms a stably folded structure which is not denatured below a temperature of 87°C, exhibits reversible thermal denaturation, and maintains its native conformation in aqueous solution containing up to approximately 8 M urea.

DNA encoding a random peptide is preferably inserted between amino acids 9 and 12 of the B1 domain at the site of a loop between two beta sheets (see Gronenborn, A.M. (1991), *supra*, for B1 structural analysis), thereby providing optimal presentation of the random peptide for binding to the target protein. The transcriptional activation domain is fused to the N-terminus of the B1 domain.

Other scaffold peptides described above can be used in a similar fashion.

In general, libraries of test vectors are constructed in which each vector contains, in the

direction of transcription, a transcriptional activation domain, a DNA sequence encoding a random amino acid sequence, and a DNA sequence encoding a scaffold peptide such as the Adh C-terminal sequence (SEQ ID NO:3) or the B1 domain of Streptococcal protein G (Gronenborn, A.M. (1991) *Science*, 253:657-661). A target vector is also constructed which contains, in the order of transcription, the Gal4 DNA binding domain linked to the coding sequence of a target protein of interest, e.g., the entire Rev coding sequence.

The relative position of the target domain and the Gal4-DNA binding domain can be reversed and still generate a functional target fusion protein. Also, as noted *supra*, other binding domains such as LexA can be used in the assay. In addition, the yeast expression vectors of the invention contain appropriate transcriptional and translational regulatory sequences and nuclear localization signal coding sequences for expression and nuclear localization of the test and target fusion proteins.

The interaction of the test random amino acid sequence and the target domain in a cell provides a functional GAL4 transcriptional activator resulting in the expression of a marker gene (e.g. *lacZ*) under the transcriptional control of a GAL4 promoter. A positive interaction is detected as blue yeast colonies on yeast culture solid medium containing X-Gal.

As the length of the random amino acid sequence increases, the number of possible random fusion proteins increases exponentially. For example, a random 20 amino acid sequence can have any one of 20 naturally occurring amino acids at each position, or $20^{20}$ different sequences. The Rev effector domain as an example of a target domain is likely to bind multiple sequences, increasing the

probability of detecting a binding reaction in the millions of independent clones screened in each assay.

Example 2: Cloning of a Protein That Binds to the Effector Domain of the HIV-1 Rev Protein

The search for a Rev effector protein was performed by initially constructing a HeLa cDNA library by standard techniques well known to those of ordinary skill in the art (HeLa S3 ATCC #CCL 2.2).

The HeLa cDNA library vector, pCF128, was constructed as follows. The *Bam*HI fragment for DB20 (Becker et al. (1991) *P.N.A.S. USA*, 88:1968-1972) containing an ADH promoter and transcription terminator expression cassette was inserted into YEPlac181 (Gietz and Sugino (1988) *Gene*, 74:527-534), which had ben digested with *Hind*III and *Eco*RI and filled in with Klenow enzyme, creating intermediate plasmid pCF124. A PCR product encoding VP16 amino acids 411-500 (Triezenberg et al. (1988), *Genes and Dev.*, 2:718) was inserted into the *Hind*III site of pCF124 creating intermediate plasmid pCF126. A *Hind*III fragment containing the polylinker of DB20 (Becker, (1991), *supra*) was inserted into the *Hind*III site of pCF126 downstream of the VP16 coding region. This generated pCF128 that contains within the yeast vector, YEPlac181, the ADH promoter, the VP16 activation domain, a polylinker, and the ADH transcription terminator which encoded the Adh C-terminal amino acids that comprise the scaffold. All cloning steps were done by standard methods (see, e.g., Ausubel et al., eds. (1987) Current Protocols in Molecular Biology, John Wiley publishers).

HeLa cDNA was prepared using the Riboclone cDNA synthesis kit (Promega Corp., Madison, WI) using random hexamer primers. Blunt end cDNA, ligated to a *Bst*XI

adapter, was inserted into the *Bst*XI sites of pCF128 to generate the test library.

The library contained approximately $4x10^6$ independent clones. This library was transformed into the yeast strain GGY1 (Gill, G. et al. (1987) *Cell* 51:121-126) containing an integrated *lacZ* reporter under GAL4 control and expressing a Gal4(1-147)-Rev fusion.

From approximately $2x10^6$ yeast transformants, 27 primary positive clones were recovered. Further screening of the primary clones eliminated false positive clones. Candidate plasmids were retransformed into GGY1 together with constructs expressing either the Gal4 DNA binding domain Gal4(1-147) alone, the Gal4 fusion Gal4-Rev, or the fusion Gal4-Rev(M10) (Table 1).

**Table 1**

| | Class of Clones | | |
|---|---|---|---|
| Constructs | I | II | III |
| Gal4(1-147) | + | - | - |
| Gal4-Rev | + | + | + |
| Gal4-RevM10 | + | + | - |
| Number of isolates: | 13 | 12 | 1 |

Of the 27 candidates, 12 library plasmids activated transcription when transformed into yeast together with either one of the Gal4-Rev fusions, but not together with Gal4 (1-147) (Class II clones). These cDNAs presumably encode proteins that bind to parts of the Rev protein that are not affected by the M10 mutation in the effector domain. Only one of the plasmids (R32) was able to discriminate between wild-type Rev and Rev-M10, causing reporter gene activation together with Gal4-Rev, but not Gal4-Rev(M10) (Class III clone). This indicates a specific binding of the protein product of

R32 to the effector domain of Rev. The R32 plasmid contains a 377 bp cDNA fragment inserted into vector pCF128 (*supra*). The expressed fusion protein contains the VP16 activation domain, the R32-encoded peptide, and the Adh scaffold peptide and was determined to be a Rev interacting protein in subsequent binding studies.

Example 3: Specificity of the Rev interaction protein-Rev Binding

To further assay the binding specificity of the isolated Rev interaction protein to Rev, the R32 plasmid was transformed into yeast together with Gal4 fusions to a number of different Rev mutants that have previously been described (Malim et al. (1991), *supra*). Binding of RIP to Rev wild-type and mutants was assayed by the method of the invention. The production of HIV-1 gtat and HIV-1 gag p24 assayed by Malim et al. (Malim et al. (1991), *supra*) is tabulated in Fig. 1 to illustrate the activity of the Rev mutants relative to wild type Rev. All of the mutants tested contain mutations in the Rev effector domain between amino acids 78 and 90, but not all of these mutations abolish Rev function (Fig. 1).

A high correlation exists between the function of these Rev mutants and their ability to bind to the RIP protein encoded in part by the original R32 sequences (Fig. 1). Mutants M20 and M25 that retain some Rev activity according to Malim et al. (1991), *supra*, bind RIP; whereas M10, and the single point mutants M27 and M29 that have no Rev activity did not show any detectable binding to RIP. Mutant M22 does not show a correlation between RIP binding and Rev function and this may be due to alteration of the amino acid and its charge at position 83 when the amino acid is changed from Leu to Asp.

Example 4: The R32 Sequence of RIP is Encoded on the Antisense Strand of a Human MCM2 Homologue

When the R32 sequence was compared with sequences in the databank (Genbank) there was no exact match, however some amino acid sequence homology was found on the antisense strand of R32 with the yeast MCM2 protein (Yan, H. et al. (1991) *Genes and Development*, 5:744-957). A human λcDNA clone (λcDNA library #936201, Stratagene, La Jolla, CA) corresponding to this region showed the R32 sequence on the antisense strand of the hMCM2 homologue. A Northern blot probed with a strand-specific RNA probe to detect messages encoding the R32 open reading frame yielded no signal above background signal; whereas the same blot hybridized with a randomly radiolabelled R32 probe (i.e. both sense and antisense strands were radiolabelled) detected one moderately abundant transcript of 4.3 kb, corresponding in size to the human MCM2 homologue. These findings indicate that the R32 sequence is not part of a cellular protein coding transcript, but rather represents a sequence selected from the 'random' sequence pool of transcripts in antisense orientation present in the cDNA library.

To confirm this hypothesis, fusion proteins expressed by various constructs were tested for Rev interaction (Fig. 2) using the two-hybrid assay. Becker et al. reported that the Adh terminator contains a cryptic promotor capable of transcribing upstream sequences (Becker, D. M. et al. (1991) *PNAS* 88:1968-1972). Fusion of the human MCM2 gene in frame with the VP16 activation domain did not result in transcriptional activation in yeast when cotransformed with the Gal4-Rev construct (Fig. 2, construct 3). Thus, the R32 fusion protein, and not the hMCM2 homologue, is responsible for Rev binding.

Surprisingly, introduction of a stop codon directly after the R32 sequence also abolished Rev binding (Fig. 2, construct #2). Closer inspection of the sequence revealed that in the original clone there was no stop codon after the cloned cDNA (R32), adding to the open reading frame 16 amino acids from polylinker sequences and 35 amino acids from the C-terminus of the yeast Adh protein present in the Adh terminator fragment used in the library test vector. This was confirmed in construct #6 (Fig. 2) where R32 sequences taken from the antisense strand of the human MCM2 open reading frame clone (without Adh sequences) again failed to interact with Rev. If a R32-Adh fusion was cloned upstream of the stop codon in clone #2 (Fig. 2), a functional protein was again obtained (Fig. 2, construct #5). A frame shift between VP16 and R32 sequences also led to an inactive construct (Fig. 2, construct #4).

Thus, the functional Rev interaction protein consisted of R32 sequences (derived from hMCM2 antisense transcript) fused to 51 amino acids from polylinker and Adh sequences (see Fig. 4). The fact that vector sequences were required for the interaction between R32-encoded amino acids and Rev indicates that R32 was not the physiological Rev target but actually represents a peptide sequence selectable when fused to a scaffold peptide domain, the Adh C-terminal sequence.

Example 5: Delineation of the Minimal Protein Sequence Required for Rev Binding

To determine the minimal amino acid sequence required for Rev binding, N- and C-terminal ExoIII deletions were performed using construct #5 (Fig. 2) which encoded a functional Rev interaction protein (containing a R32-Adh scaffold fusion). Plasmid pools of different time points from unidirectional ExoIII deletion

experiments were cotransformed into yeast with the Gal4-Rev target vector. Plasmids were rescued from positive yeast colonies. The shortest functional peptide recovered from an N-terminal deletion series had 203 of the 377 bp of the original R32 sequence deleted. The construct was transformed into yeast strains expressing Gal4(1-147), Gal4-Rev, and Gal4-RevM10 to assess binding specificity. The N-terminal deletion peptide not only retained full discrimination between Rev and RevM10 but, judging from the levels of $\beta$-Gal produced by these strains, the interaction between the N-terminal deleted Rev interaction peptide and Rev was stronger than the interaction detected with the original Rev interaction protein isolate.

To determine the number of C-terminal amino acids that could be deleted from the original Rev interaction protein isolate without loss of Rev binding activity, a similar strategy of ExoIII deletions was used. These manipulations were again performed with clone #5 (Fig. 2). Surprisingly, the active constructs from late deletion time points all showed the same structure: the polylinker and the Adh domain upstream of the stop codon had been deleted. Also, all of the recovered active clones no longer had the stop codon and were now in frame with the downstream Adh domain. Since, in these experiments, no active clones were recovered which lacked the downstream Adh domain fused in frame with upstream R32 sequences, conclusions regarding a minimal active protein can be drawn. Only one amino acid can be deleted from the C-terminus of the R32 sequence, and the polylinker sequences are dispensable, whereas all 35 vector-encoded amino acids (SEQ ID NO:3) originating from the yeast Adh protein are necessary for an interaction with Rev. The minimal amino acid sequence, RIP (Rev interacting protein, SEQ ID NO:2), is shown in Fig. 4.

The original R32 clone was 377 nucleotide base pairs in length and encoded 125 amino acids. The two-hybrid library test vector polylinker added 16 amino acids and the last 35 amino acids were contributed from the Adh terminator in the expression cassette. Deletion analysis showed that 67 of the unique 125 amino acids as well as the polylinker sequences could be deleted from the construct without losing Rev binding ability. The minimal protein sequence therefore consists of 58 amino acids encoded by the hMCM2 antisense fragment fused to the Adh scaffold of 35 amino acids (SEQ ID NO:3) of the Adh C-terminus.

Example 6: RIP Expression Inhibits HIV Viral Protein Expression

Rev is essential for HIV replication in human cells and the Rev effector domain is required for its function (Malim, M. et al. (1989), *supra*). The RIP peptide, containing a portion of a cDNA isolated from the random pool of normally untranslated messages fused to the Adh-derived scaffold, was tested to determine if its Rev binding ability correlated with an ability to inhibit Rev function.

For these tests, the HIV-1 infectious molecular clone, pNL4-3 (Adachi, A. et al. (1986) *J. Virol.* 59:284-291), which directs the expression of infectious HIV-1 virions following transfection into a variety of cell types, was used as the source of HIV in human cells. Preliminary experiments were performed in HeLa cells (HeLa S3 ATCC #CCL 2.2). HeLa cells were grown in monolayers in DMEM media (GIBCO, Grand Island, NY) supplemented with 10% fetal bovine serum. Cells were cotransfected with pNL4-3 and either a CMV promoter-driven expression vector (pcDNA1, Invitrogen, San Diego, CA) or the identical expression vector

containing the RIP cDNA (SEQ ID NO:1) (pcRIP) cloned into the *Hin*dIII and *Eco*RI sites of pcDNA1.

Transfections were performed on monolayer cells by introducing 20 μg of total plasmid DNA by the calcium phosphate procedure (Chen C. et al. (1988) *Biotechniques*, 6:632-638) using 5 μg of LTR-CAT (chloramphenicol acetyltransferase) reporter construct, 5 to 15 μg of test plasmid DNA, and carrier DNA. Suspension cultures were transfected with DEAE-dextran (Margolis et al., (1993) *J. Virology*, 192:370-374) using 4 μg of pNL4-3 and 4-12 μg of test plasmid DNA. To assay HIV-1 replication, aliquots of culture media were sampled for detection of HIV-1 gag p24 protein (Veronese, et al. (1985) *P.N.A.S. USA*, 82:5199-5202). Assays were performed using an antigen-capture ELISA kit (Coulter, Hialeah, FL).

Inhibition of HIV-1 production increased with cotransfection of increasing amounts of RIP expression vector relative to a constant amount of pNL4-3 HIV-1 molecular clone (Fig. 5). Expression of a co-transfected HIV LTR-driven CAT reporter gene in these cells was unaffected by RIP expression, demonstrating that inhibition of virus production was not due to transcriptional effects, or non-specific toxicity.

Example 7: RIP Inhibits HIV Production in Lymphocytes

As lymphocytes are a primary target of HIV-1 infection *in vivo*, further experiments were performed in lymphocyte cell lines. In CEM cells, cotransfection of RIP-encoding DNA at a transfection ratio of 2:1 (4 μg pNL4-3 and 8 μg pcRIP) resulted in inhibition of HIV-1 antigen production up to twenty-fold greater than that mediated by cotransfection of the dominant negative Rev mutant M10 (Fig. 6A). Previous studies demonstrating potent inhibition by Rev M10 in similar cotransfection studies have used transfection ratios of 10:1 (Malim et

al. (1989), *supra*). Thus, RIP is a more potent inhibitor requiring a two-fold excess (2:1) of RIP to inhibit whereas a 10-fold excess (10:1) of M10 is required for inhibition of HIV-1 antigen production. Cell viability and cell growth were unaffected by RIP expression as tested using well known techniques of cell counting and trypan blue assay. Cotransfection of a secreted alkaline phosphatase reporter gene (pCMU/SEAP) (Tropix Inc., Bedford, MA; Cullen, B. et al. (1992) *Meth. Enzyme*, 216:362-368) showed that RIP did not impair the ability of these cells to express transfected genes (Fig. 6B). Decreased expression of alkaline phosphatase at later time points in cells cotransfected with Rev M10 or pcDNA1 may reflect cell killing by HIV-1. Similar cotransfection experiments performed in H9 lymphocytes also demonstrated the ability of RIP to inhibit virion production in a dose dependent manner (Fig. 7).

T cell lines, CEM and H9 were maintained in suspension culture in RPMI 1640 media (ICN Biochemicals, Cosa Mesa, CA) supplemented with 10% fetal bovine serum (GIBCO, Grand Island, NY).

Example 8: RIP Inhibition of HIV Is Not Due to Antisense-Mediated Inhibition of MCM2 Expression

Since most of the RIP peptide has been shown here to be encoded by the antisense strand of a human MCM2 homologue, an important control is to show that the observed inhibitory affect on HIV-1 production was not due to antisense transcript-mediated inhibition of endogenous hMCM2. The yeast MCM2 protein has been shown to be involved in plasmid maintenance (Yan, H. et al. (1991) *Genes and Development*, 5:944-957), and is thought to be involved in the initiation of DNA replication (Yan H. et al. (1991), *supra*). A frame shift mutation was generated directly after the ATG in the RIP

expressing vector. This plasmid therefore retains all of the antisense sequence, however it is not able to express a Rev binding peptide. Fig. 7 shows that such a construct (pc ΔRIP) has no antiviral activity.

Example 9: Assay to Detect the Presence and/or Quantity of Rev in a Cell Lysate Using RIP and Anti-RIP Antibody

The RIP of the invention as well as anti-RIP antibodies are useful in an assay for the presence of Rev in cells of a patient. Thus, the assay is useful as a diagnostic tool to determine whether cells (e.g. lymphocytes) of a patient contain the HIV virus.

To determine the presence of actively replicating HIV in a cell population, lymphocytes of a human suspected of being infected with HIV-1 are collected, suspended in an appropriate culture medium to induce cellular and viral replication and induce expression of viral proteins such as Rev. The cells are then lysed and the supernatant containing Rev is contacted with RIP immobilized on the surface of the well of a standard cell culture dish as shown in Fig. 8 (Steps A and B). After specific binding of Rev to the immobilized RIP, the remaining, unbound proteins of the lysate are removed by washing (Fig. 8, Step C). Next, an anti-RIP antibody is contacted with the proteins in the cell culture well (Fig. 8, Step D). The anti-RIP antibody binds to the remaining free RIP peptides. When Rev protein is not present in the lysate, the signal is high. When the lysate contains Rev, the signal is reduced due to a decrease in the number of accessible RIP proteins in the well.

The anti-RIP antibody can be directly labelled with a detectable tag such as radiolabelling with $^{125}I$ or a covalently bound fluorescent moiety, e.g., fluorescein. Further, unlabeled anti-RIP antibodies can be used, and

the signal generated by the secondary binding of a labeled antibody to anti-RIP antibody. Further, it is a feature of the invention that, under the conditions of the assay, the anti-RIP antibody binds RIP with reduced affinity relative to the binding of Rev to RIP. Thus, the Rev protein remains bound to RIP during the assay for reliable qualitative or quantitative determination of the presence of Rev in cells.

The assay is also useful in screening candidate agents that inhibit viral production by inhibiting the expression of essential viral proteins such as Rev. According to this embodiment of the invention, a cell culture containing actively replicating HIV is contacted with a candidate agent in an appropriate culture medium. The cells are assayed as above for Rev protein. Successful agents are those which reduce the amount of Rev protein by at least approximately 90% in the treated cells relative to untreated control cells.

Standard protocols are followed to carry out an assay method of the invention. For example, after bringing the immobilized RIP into contact with a cell lysate, the surface is washed to remove unbound material. Further, standard blocking protocols are generally used prior to contacting the washed surface with labeled antibodies. An alternative to using cells infected with HIV is to provide acutely virus infected cells. For example, lymphocytes are obtained from uninfected persons and infected with a defined amount of HIV. The cells are then treated to remove input virus and cultured.

Isolation of RIP of other selected random peptide-scaffold fusion proteins and production of antibodies to them are performed using standard techniques well known to those of ordinary skill in the art. The selected random peptide-scaffold peptide of the invention can be produced by first transforming a suitable host cell with

the nucleotide sequence encoding the random peptide-scaffold fusion protein cloned into a suitable expression vehicle followed by expression of the desired protein or polypeptide.

Those of ordinary skill in the field of molecular biology will understand that any of a wide variety of expression systems can be used to provide the protein or polypeptide. The precise host cell used is not critical to the invention. The polypeptide can be produced in a prokaryotic host (e.g., *E. coli*) or in a eukaryotic host (e.g., *Saccharomyces cerevisiae*). The method of transformation of the cells and the choice of expression vehicle will depend on the host system selected. Methods described herein provide sufficient guidance to successfully carry out the production, purification and identification of RIP or other selected random peptide-scaffold fusion protein.

Once the RIP or other selected random peptide-scaffold fusion protein is expressed, it is isolated, e.g., using immunoaffinity chromatography. In one example, an anti-RIP antibody can be attached to a column and used to isolate RIP. Lysis and fractionation of RIP-containing host cells prior to affinity chromatography can be performed by standard methods. Once isolated, the recombinant protein can, if desired, be further purified, e.g., by high performance liquid chromatography (see, e.g., Fisher, Laboratory Techniques In Biochemistry and Molecular Biology, eds., Work and Burdon, Elsevier, (1980)). RIP or other peptides can also be produced by chemical synthesis by standard solution or solid phase peptide synthesis techniques.

Substantially pure RIP or other random peptide-scaffold fusion protein can be used to raise antibodies. Antibodies directed to the polypeptide of interest are produced as follows. Peptides corresponding to all or

part of the polypeptide of interest are produced using a peptide synthesizer by standard techniques, or are isolated and purified as described above. The peptides are coupled to KLH with m-maleimide benzoic acid N-hydroxysuccinimide ester. The KLH-peptide is mixed with Freund's adjuvant and injected into animals, e.g. guinea pigs or goats, to produce polyclonal antibodies.

Monoclonal antibodies can be prepared using the polypeptide of interest described above and standard hybridoma technology (see, e.g., Kohler et al., Nature (1975) 256:495; Kohler et al., Eur. J. Immunol. (1976) 6:292; Kohler et al., Eur. J. Immunol. (1976) 6:511; Hammerling et al., in Monoclonal Antibodies and T Cell Hybridomas, Elsevier, NY, (1981), which are incorporated herein by reference). Antibodies are purified by peptide antigen affinity chromatography.

Once produced, antibodies to RIP, for example, are tested for specific RIP binding by Western blot or immunoprecipitation analysis by standard techniques.

## Uses

The method of isolating a random peptide-scaffold peptide fusion protein is useful to discover peptides that bind specifically to selected target proteins. As described above, the binding of the random peptide is enhanced by the 3-dimensional structure assumed in the context of the stably folded scaffold peptide.

The fused scaffold structure is also useful to determine the 3-dimensional structure of the random peptide such that small organic molecules can be designed to mimic some of its structural features. These small organic molecules are useful as potential therapeutic agents where binding to the target protein affects its function, e.g., inhibition of Rev protein.

The isolated random peptide-scaffold fusion protein RIP is useful as an inhibitor of HIV production in cells, such as human lymphocyte cells. RIP and antibodies to RIP are useful in an assay for the presence of Rev in a lysate of cells suspected of being infected with HIV.

RIP can be used to treat or prevent HIV infection by introducing a DNA sequence (SEQ ID NO:1) encoding RIP (SEQ ID NO:2), or Rev interacting analogues of the original R32-scaffold fusion, into the genome of a human cell infected with, or likely to be infected with, HIV. RIP is expressed in the cell at sufficient concentration within the cell to inhibit Rev function and, thereby inhibit HIV replication.

Cells which are transfected with RIP-encoding DNA sequence include cells that can be infected by HIV such as $CD4^+$ expressing cells. In one method, bone marrow stem cells from an infected patient are removed from the patient and then infected with a an appropriate viral vector containing RIP-encoding DNA sequence by standard gene therapy techniques well known in the art (see e.g., Morsy, M.A. et al. (1993) *JAMA* 270:2338-2345). The patient's remaining bone marrow is preferably destroyed by irradiation or chemical ablation to eliminate any HIV-infected stem cells that do not express RIP. The transfected stem cells are reintroduced into the host from which they were isolated (autologous transplantation) by injection into a vein of the host (see, Biron, P. et al. (1993) in Autologous Bone Marrow Transplantation: Proceedings of the First International Symposium, Dicke, K.A. et al. (eds.), p. 203 for relevant techniques).

The transfected stem cells are allowed to repopulate the bone marrow of the patient and provide $CD4^+$ cells expressing RIP which inhibits HIV replication

in these cells when infected with the virus. This results in simultaneously treating and protecting the host from disease resulting from HIV infection and replication. Allogenic transplantation of stem cells expressing RIP from an uninfected host is another embodiment of the invention and provides the same protection to transfected cells and their progeny from subsequent HIV replication.

Another use for a random peptide isolated by the method of the invention involves analyzing the structure of the random peptide in the context of the scaffold peptide for designing small molecule therapeutic agents. First, DNA encoding a random peptide is inserted into the coding sequence of the scaffold peptide, e.g., the B1 domain of G protein. In this context, the random peptide is displayed on the surface of a stable protein and therefore is less likely to be degraded. Also the structural constraints of the scaffold may maintain the peptide in an active conformation. The structure of the random peptide necessary to interact with the target protein is deduced by comparing the 3-dimensional structure of the scaffold peptide domain with and without the inserted random peptide sequence, e.g., by using NMR analysis (Heinz et al. (1994) *J. Mol. Biol.* 236:869-886). Such an analysis would not be possible without the structural stabilizing effects of the scaffold peptide domain in combination with the previously solved NMR analysis of the scaffold structure.

Peptides that are not constrained by a scaffold peptide domain may not have a defined structure in solution. They assume a specific structure only when binding to their target protein, i.e., an induced fit. Structural analysis of the random peptide inserted in a scaffold peptide domain obviates the need for analysis of the target protein-random peptide complex.

Three-dimensional information of an effective binding structure provides useful information for designing small organic molecules that can also bind the target protein. For example, a small organic molecule which mimics an aspect of the 3-dimensional structure of RIP can be designed and tested for its ability to inhibit HIV production in cells.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: University of Massachusetts Medical Center

(ii) TITLE OF INVENTION: METHODS OF SELECTING A RANDOM PEPTIDE THAT BINDS TO A TARGET PROTEIN

(iii) NUMBER OF SEQUENCES: 7

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Fish & Richardson
(B) STREET: 225 Franklin Street, Suite 3100
(C) CITY: Boston
(D) STATE: MA
(E) COUNTRY: USA
(F) ZIP: 02110-2804

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/381,800
(B) FILING DATE: 01-FEB-1995

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Clark, Paul T.
(B) REGISTRATION NUMBER: 30,162
(C) REFERENCE/DOCKET NUMBER: 04020/062WO1

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (617)542-5070
(B) TELEFAX: (617)542-8906
(C) TELEX: 200154

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 330 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ACGTACCTTG TGCTTGCCAC CTGGGTTTTT GGGCTCCCCT CCGAACAGGG CCAGAGCCAG 60

46CCTCTCTTG ATGTCTTCAT GACCAGAGAT GGAAGGAGCA ATGCTGGCAA AGATCTTCTC 120

TCCGATCTGC TGATCCTTGG AGAGGCTAGT GATCATCTTC ACATCTTCAT CGGTCCTGGC 180

GAGCACACTG GCGGCCGCTC GAAGCTTTGG ACTTCTTCGC CATTGGTCAA GTCTCCAATC 240

AAGGTTGTCG GCTTGTCTAC CTTGCCAGAA ATTTACGAAA AGATGGAAAA GGGTCAAATC 300

GTTAGATACG TTGTTGACAC TTCTAAATAA 330

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 109 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Thr Tyr Leu Val Leu Ala Thr Trp Val Phe Gly Leu Pro Ser Glu Gln
1 5 10 15

Gly Gln Ser Gln Ala Ser Leu Asp Val Phe Met Thr Arg Asp Gly Arg
20 25 30

Ser Asn Ala Gly Lys Asp Leu Leu Ser Asp Leu Leu Ile Leu Gly Glu
35 40 45

Ala Ser Asp His Leu His Ile Phe Ile Gly Pro Gly Glu His Thr Gly
50 55 60

Gly Arg Ser Lys Leu Trp Thr Ser Ser Pro Leu Val Lys Ser Pro Ile
65 70 75 80

Lys Val Val Gly Leu Ser Thr Leu Pro Glu Ile Tyr Glu Lys Met Glu
85 90 95

Lys Gly Gln Ile Val Arg Tyr Val Val Asp Thr Ser Lys
100 105

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Leu Val Lys Ser Pro Ile Lys Val Val Gly Leu Ser Thr Leu Pro Glu
1 5 10 15

Ile Tyr Glu Lys Met Glu Lys Gly Gln Ile Val Arg Tyr Val Val Asp
20 25 30

Thr Ser Lys
35

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 175 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

ACGTACCTTG TGCTTGCCAC CTGGGTTTTT GGGCTCCCCT CCGAACAGGG CCAGAGCCAG 60

GCCTCTCTTG ATGTCTTCAT GACCAGAGAT GGAAGGAGCA ATGCTGGCAA AGATCTTCTC 120

TCCGATCTGC TGATCCTTGG AGAGGCTAGT GATCATCTTC ACATCTTCAT CGGTC 175

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 58 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Thr Tyr Leu Val Leu Ala Thr Trp Val Phe Gly Leu Pro Ser Glu Gln
1 5 10 15

Gly Gln Ser Gln Ala Ser Leu Asp Val Phe Met Thr Arg Asp Gly Arg
20 25 30

Ser Asn Ala Gly Lys Asp Leu Leu Ser Asp Leu Leu Ile Leu Gly Glu
35 40 45

Ala Ser Asp His Leu His Ile Phe Ile Gly
50 55

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 108 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TTGGTCAAGT CTCCAATCAA GGTTGTCGGC TTGTCTACCT TGCCAGAAAT TTACGAAAAG 60

ATGGAAAAGG GTCAAATCGT TAGATACGTT GTTGACACTT CTAAATAA 108

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Pro Val Pro Leu Gln Leu Pro Pro Leu Glu Arg Leu Thr Leu Asp Cys
1 5 10 15

Asn Glu Asp Cys Gly Thr Ser
20

## Other Embodiments

Other embodiments are within the following claims.

What is claimed is:

## Claims

1. A method of isolating a random peptide that binds a predetermined target protein, said method comprising:

a) providing a library of nucleotide sequences encoding random peptides, each of said sequences encoding a candidate random peptide;

b) cloning a sequence from said library into a first yeast two-hybrid expression vector to make a library test vector expressing a random peptide-scaffold fusion protein comprising

i) an N-terminal transcriptional activation domain,

ii) said candidate random peptide, and

iii) a scaffold peptide domain;

c) cloning a nucleotide sequence encoding a target peptide domain into a second expression vector to make a target vector expressing a target fusion protein comprising

i) an N-terminal DNA binding domain, and

ii) a C-terminal target binding domain;

d) introducing said library test vector and said target vector into a yeast cell wherein said cell contains a marker gene expressed when the test fusion protein binds to the target fusion protein; and

e) isolating said random peptide-scaffold fusion protein from a cell in which said marker gene is expressed.

2. A method of claim 1, wherein said random peptide sequence is inserted within a scaffold peptide domain.

3. A method of claim 1, wherein said scaffold peptide domain is a peptide comprising SEQ ID NO:3.

4. A method of claim 1, wherein said scaffold peptide domain is a peptide comprising the B1 domain of G protein.

5. A method of claim 1, wherein said scaffold peptide domain is a peptide comprising a zinc finger domain, the B domain of protein A of *S. aureus*, the tenth type III module of fibronectin, or cyclophilin.

6. Substantially pure DNA of SEQ ID NO:1, or degenerate variants thereof, encoding an HIV-1 Rev binding peptide.

7. A substantially pure peptide comprising an amino acid sequence substantially identical to SEQ ID NO:2, which peptide binds to HIV-1 Rev protein.

8. Substantially pure DNA of SEQ ID NO:4, or degenerate variants thereof.

9. A substantially pure peptide comprising an amino acid sequence substantially identical to SEQ ID NO:5.

10. An antibody that binds specifically to the peptide of claim 7.

11. A vector containing the DNA of claim 6 operably linked to transcriptional and translational regulatory sequences, wherein said vector expresses said DNA as an HIV-1 Rev binding peptide in a mammalian cell.

12. A method of inhibiting HIV-1 replication in a mammalian cell, said method comprising:

a) introducing a vector of claim 11 into a mammalian cell infected with HIV-1; and

b) expressing said HIV-1 Rev binding peptide in said cell at a concentration within said cell that inhibits HIV-1 replication such that virion production is reduced within said cell.

13. A method of treating a human for HIV-1 infection, said method comprising:

a) providing a vector of claim 11 in a pharmaceutically acceptable carrier;

b) introducing said vector into a human cell such that said DNA sequence is incorporated into the genome of said cell and said HIV-1 Rev binding peptide is expressed in sufficient concentration in said cell to inhibit virion production; and

c) administering said vector-containing cell into said human such that said cell and its progeny are viable in said human, wherein said cell and its progeny inhibit virion production.

14. A method of detecting human immunodeficiency virus (HIV) in a sample of human $CD4^+$ cells, said method comprising:

a) providing a sample of human $CD4^+$ cells;

b) inducing replication of any HIV in said cells;

c) lysing said cells and collecting a cell supernatant containing cellular and viral proteins;

d) contacting said cell supernatant with a peptide of claim 7, and allowing formation of peptide-Rev protein complexes; and

e) detecting said peptide-Rev protein complexes as an indication of HIV in the cell sample.

15. A method of screening for candidate agents that inhibit production of HIV by inhibiting production of Rev protein in a human cell, said method comprising:

a) providing a population of human cells containing actively replicating HIV in culture medium;

b) contacting said population of cells with a candidate agent to inhibit Rev production;

c) lysing said cells and collecting a cell supernatant containing cellular and viral proteins;

d) contacting said cell supernatant with the peptide of claim 7 and allowing formation of peptide-Rev protein complexes;

e) detecting said peptide-Rev protein complexes as an indication of Rev protein production in the cell sample;

f) repeating said steps a), c), d), and e) to establish a control sample; and

g) comparing said Rev protein production in the presence of said candidate agent with Rev protein production in said control sample to screen candidate agents that inhibit production of HIV.

SUBSTITUTE SHEET (RULE 26)

| | aa70 | | | | | | | | | | | | | | | | | | | | | | 92 | gtat | p24 (% of wt) | binding to RIP in yeast |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| wtREV: | P | V | P | L | Q | L | P | P | L | E | R | L | T | L | D | C | N | E | D | C | G | T | S | + + | 100 | ++ |
| M 10: | - | - | - | - | - | - | - | - | D | L | - | - | - | - | - | - | - | - | - | - | - | - | - | - - | <1 | -- |
| M 20: | - | - | - | - | - | - | - | - | - | D | L | - | - | - | - | - | - | - | - | - | - | - | - | + + | 50 | ++ |
| M 22: | - | - | - | - | - | - | - | - | - | - | - | - | - | D | L | - | - | - | - | - | - | - | - | - - | <1 | ++ |
| M 25: | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | | - | - | E | L | D | - | - | + + | 70 | +++ |
| M 27: | - | - | - | - | - | - | - | - | A | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - - | <1 | -- |
| M 29: | - | - | - | - | - | - | - | - | - | - | - | - | - | A | - | - | - | - | - | - | - | - | - | - - | <1 | -- |





FIG. 1

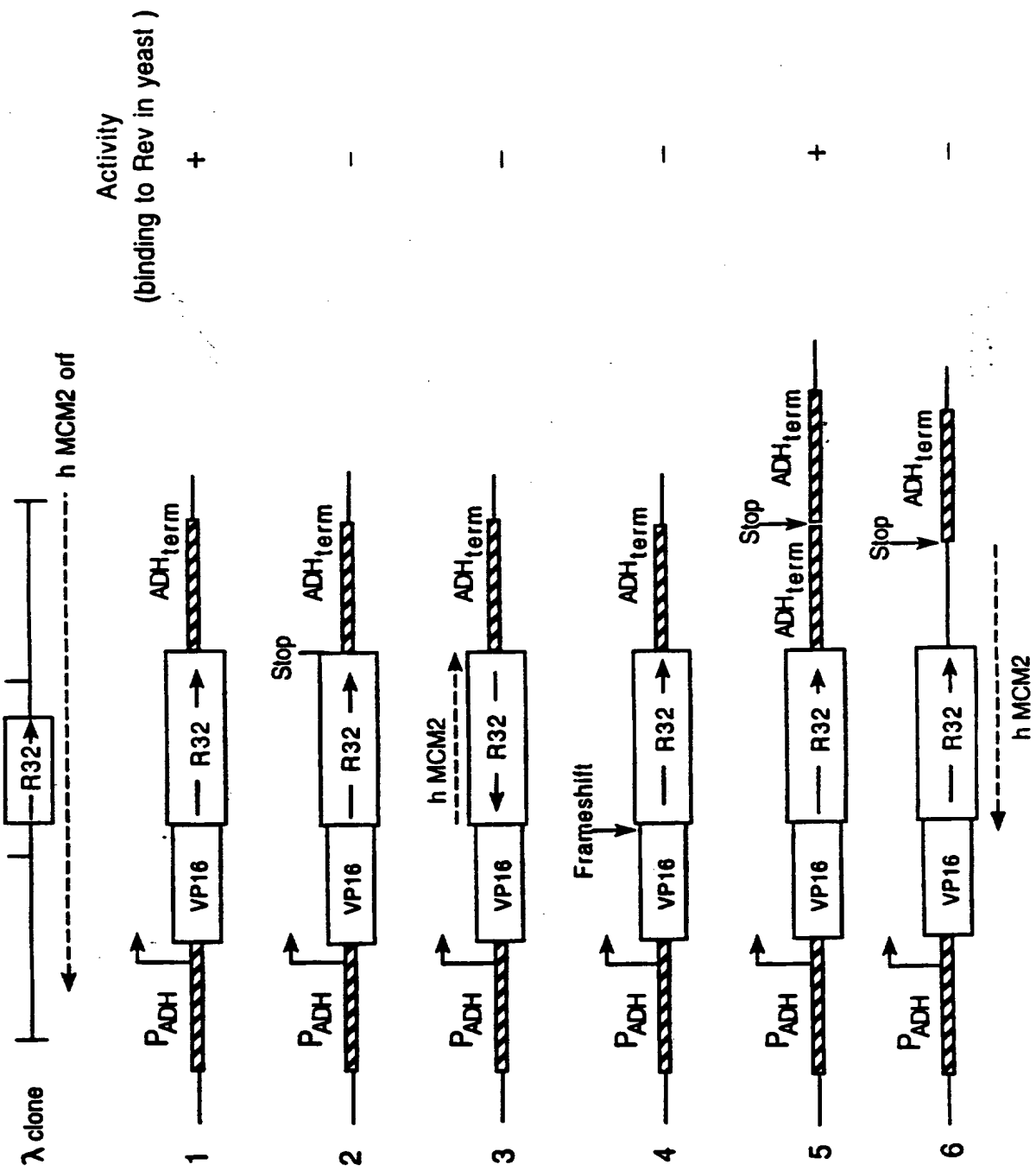
Activity
(binding to Rev in yeast )
+
-
-
-
+
-
λ clone
R32
h MCM2 orf
1
2
3
4
5
6
P ADH
VP16
R32
ADH term
Stop
h MCM2
Frameshift

FIG. 2

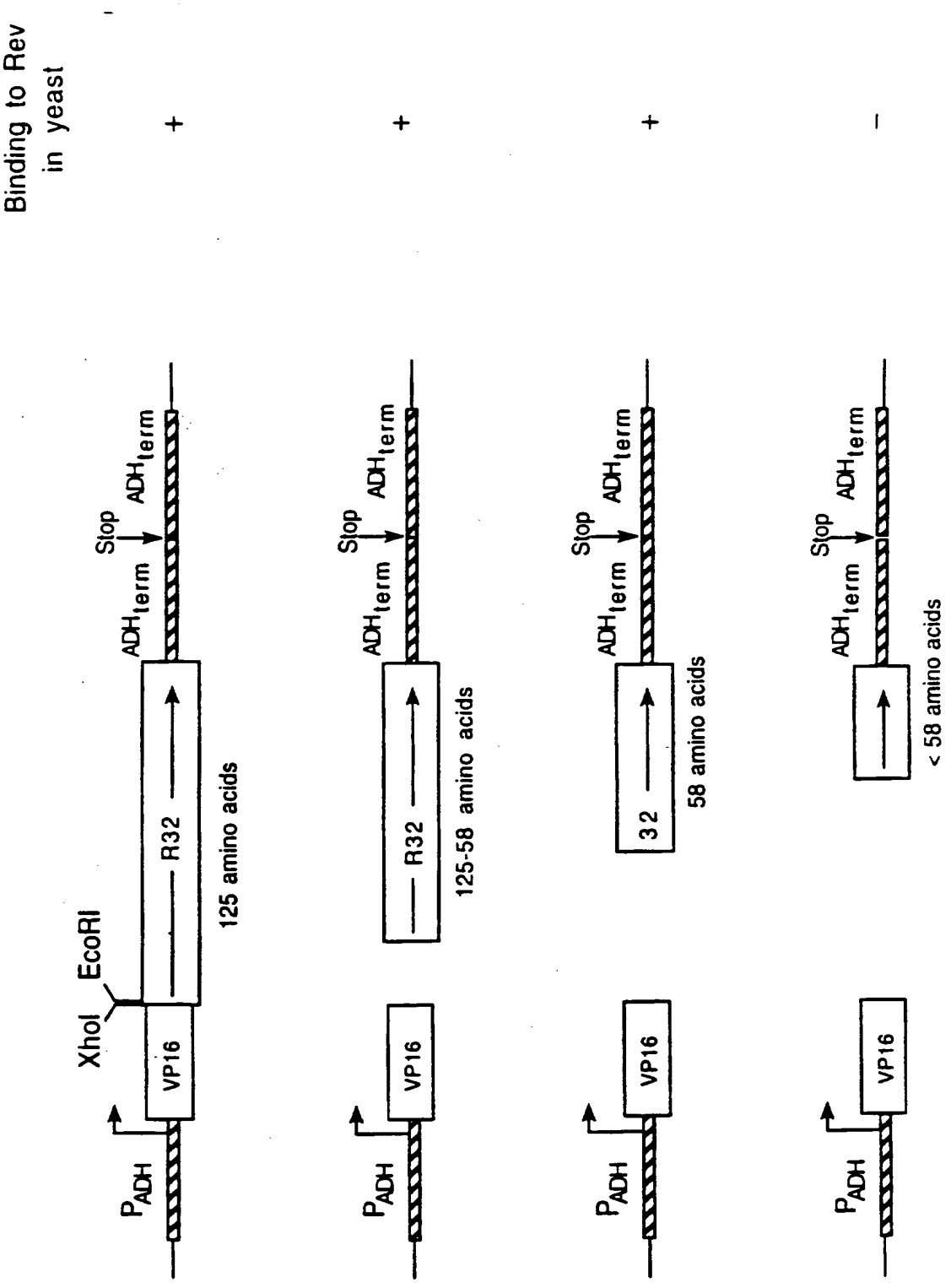


FIG. 3A

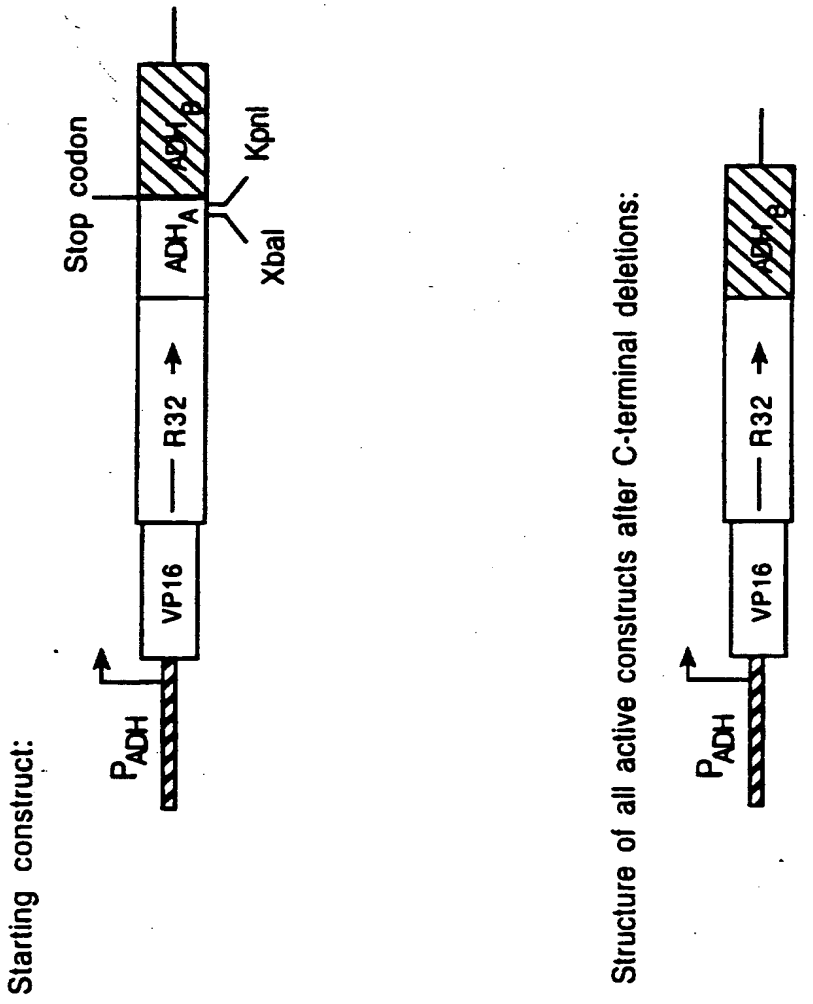
Starting construct:
$P_{ADH}$
VP16
R32
$ADH_A$
$ADH_B$
Stop codon
KpnI
XbaI
Structure of all active constructs after C-terminal deletions:
$P_{ADH}$
VP16
R32
$ADH_B$

FIG. 3B

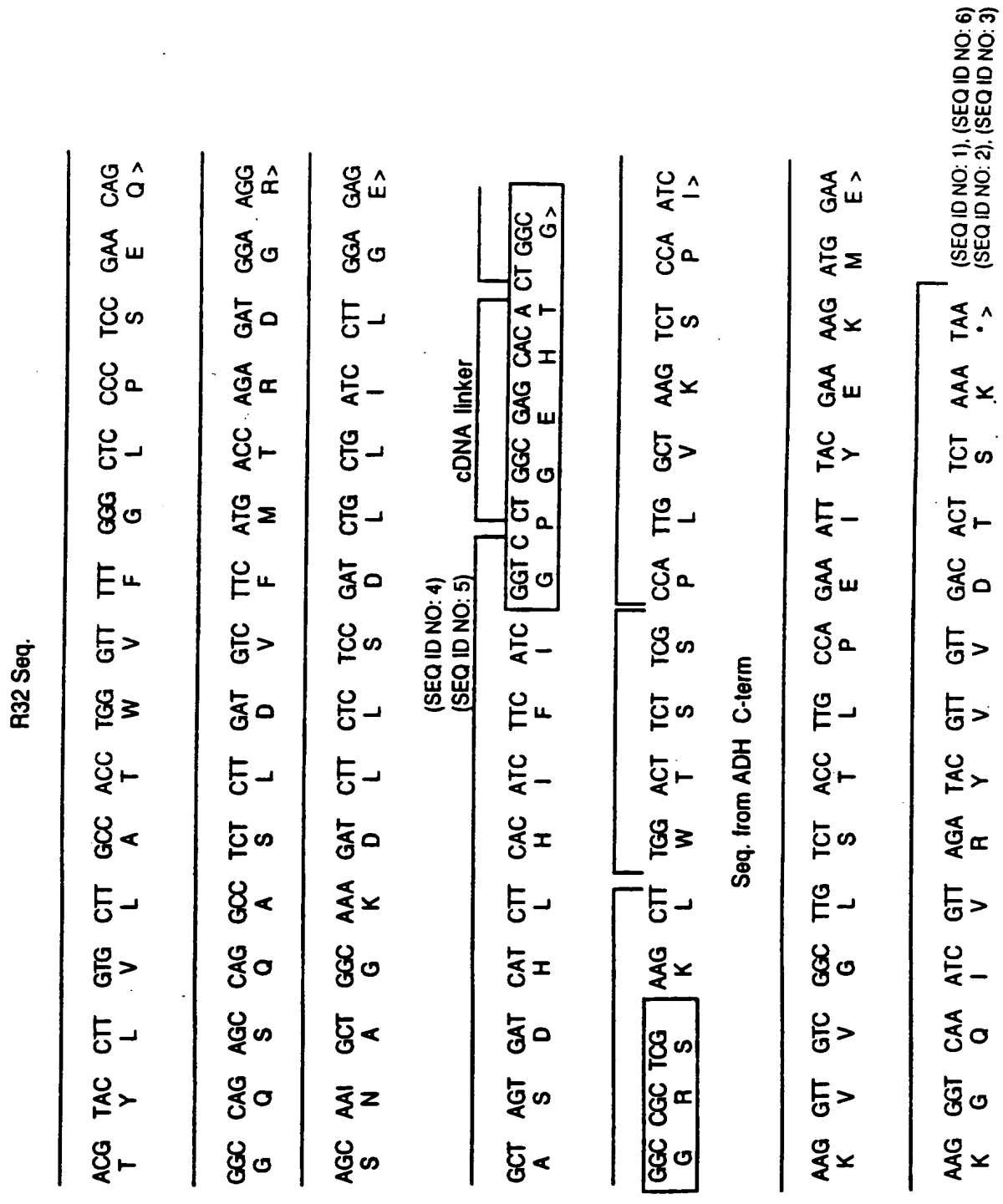


FIG. 4

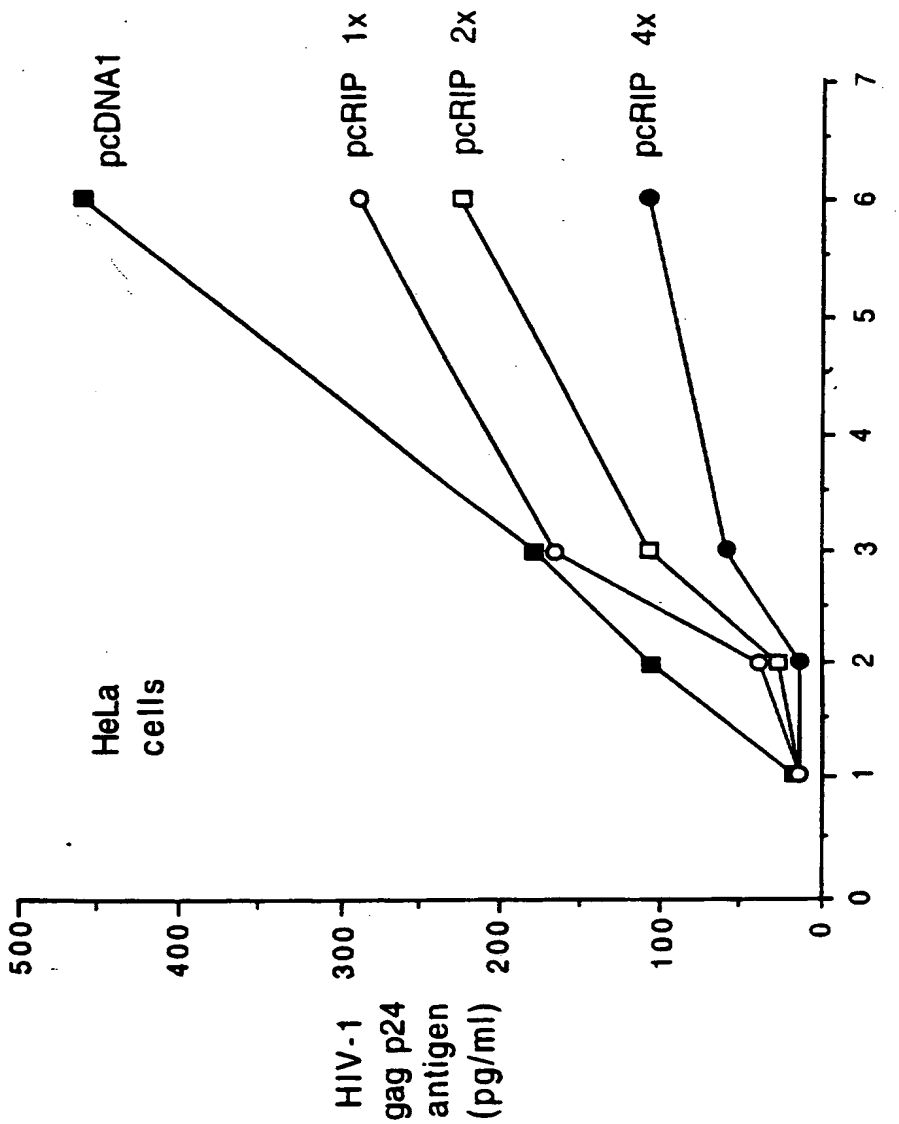
pcDNA1
pcRIP 1x
pcRIP 2x
pcRIP 4x
HeLa cells
HIV-1 gag p24 antigen (pg/ml)
500
400
300
200
100
0
0
1
2
3
4
5
6
7
days in culture

FIG. 5

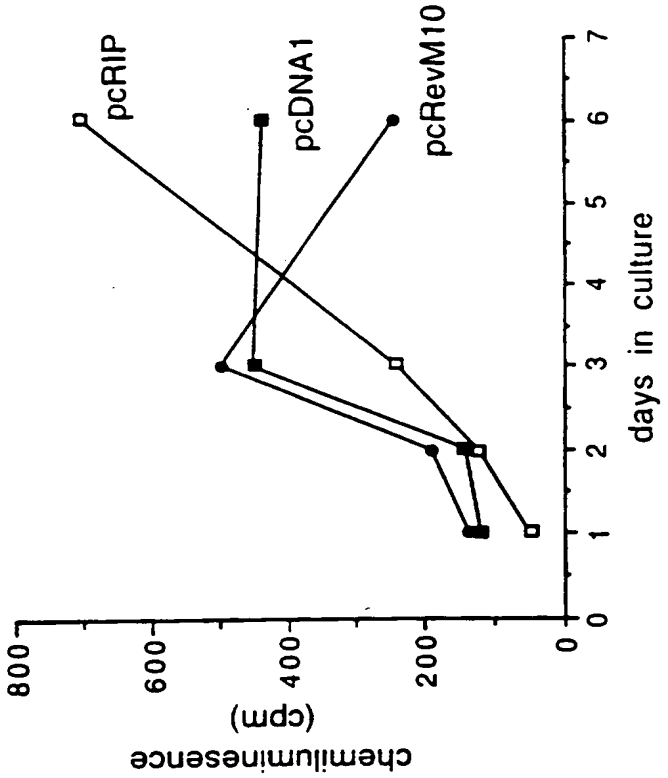


FIG. 6B

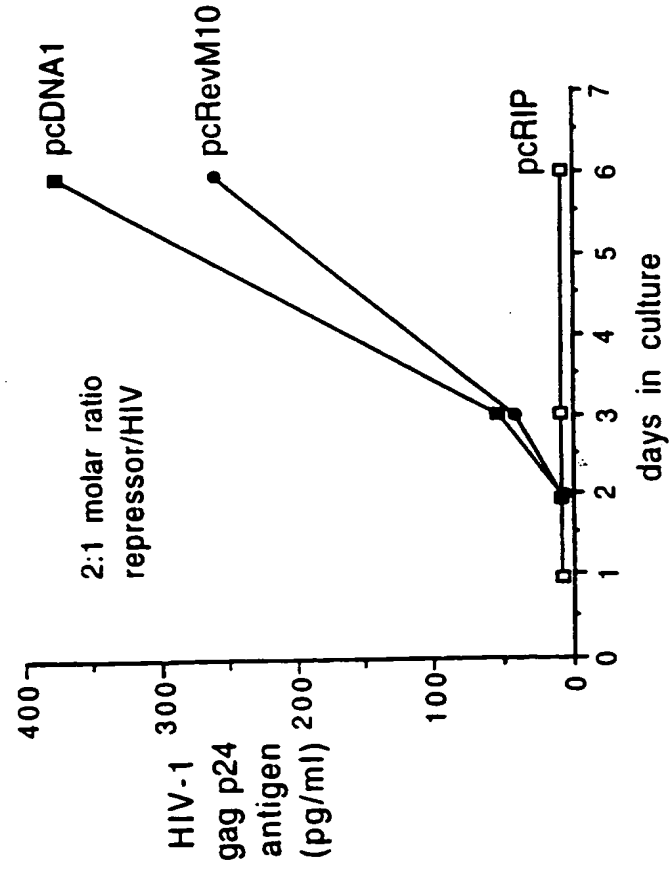


FIG. 6A

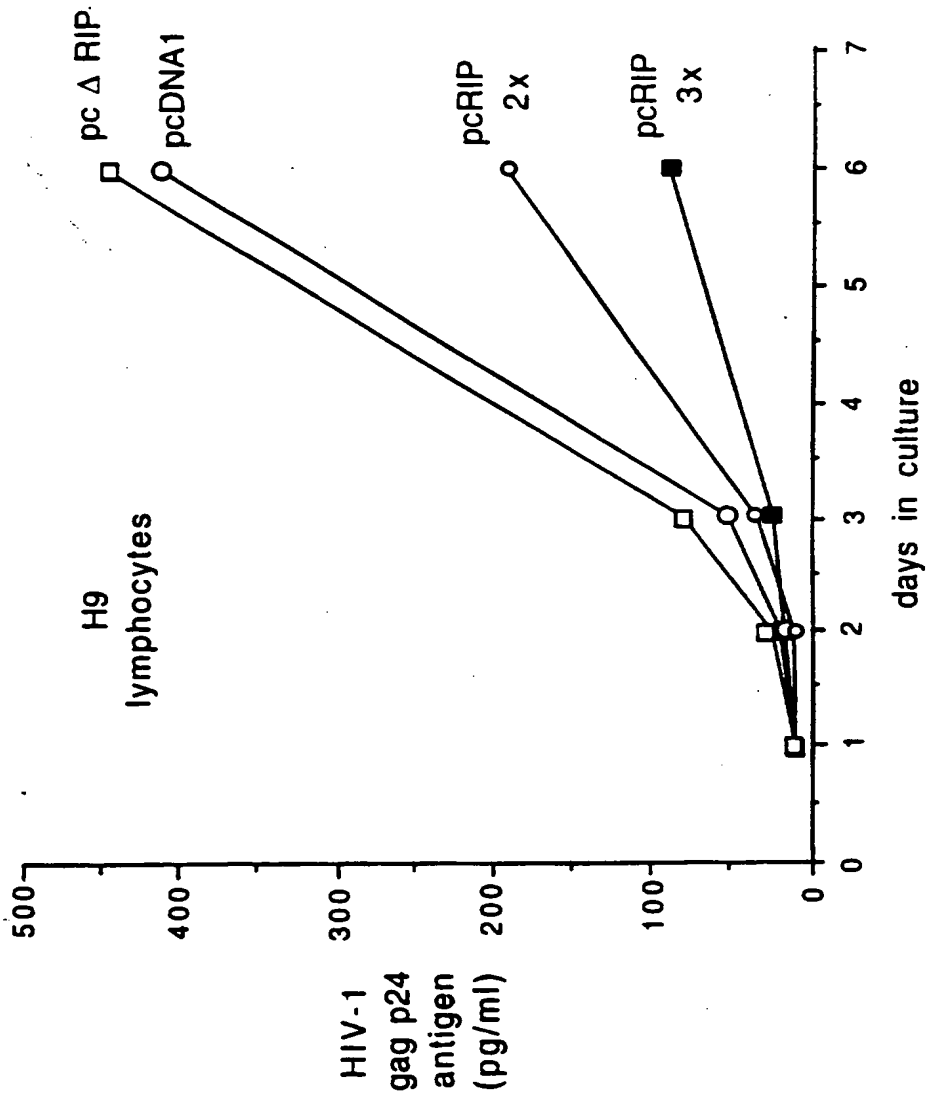
pc Δ RIP
pcDNA1
pcRIP 2x
pcRIP 3x
H9 lymphocytes
HIV-1 gag p24 antigen (pg/ml)
500
400
300
200
100
0
0
1
2
3
4
5
6
7
days in culture

FIG. 7

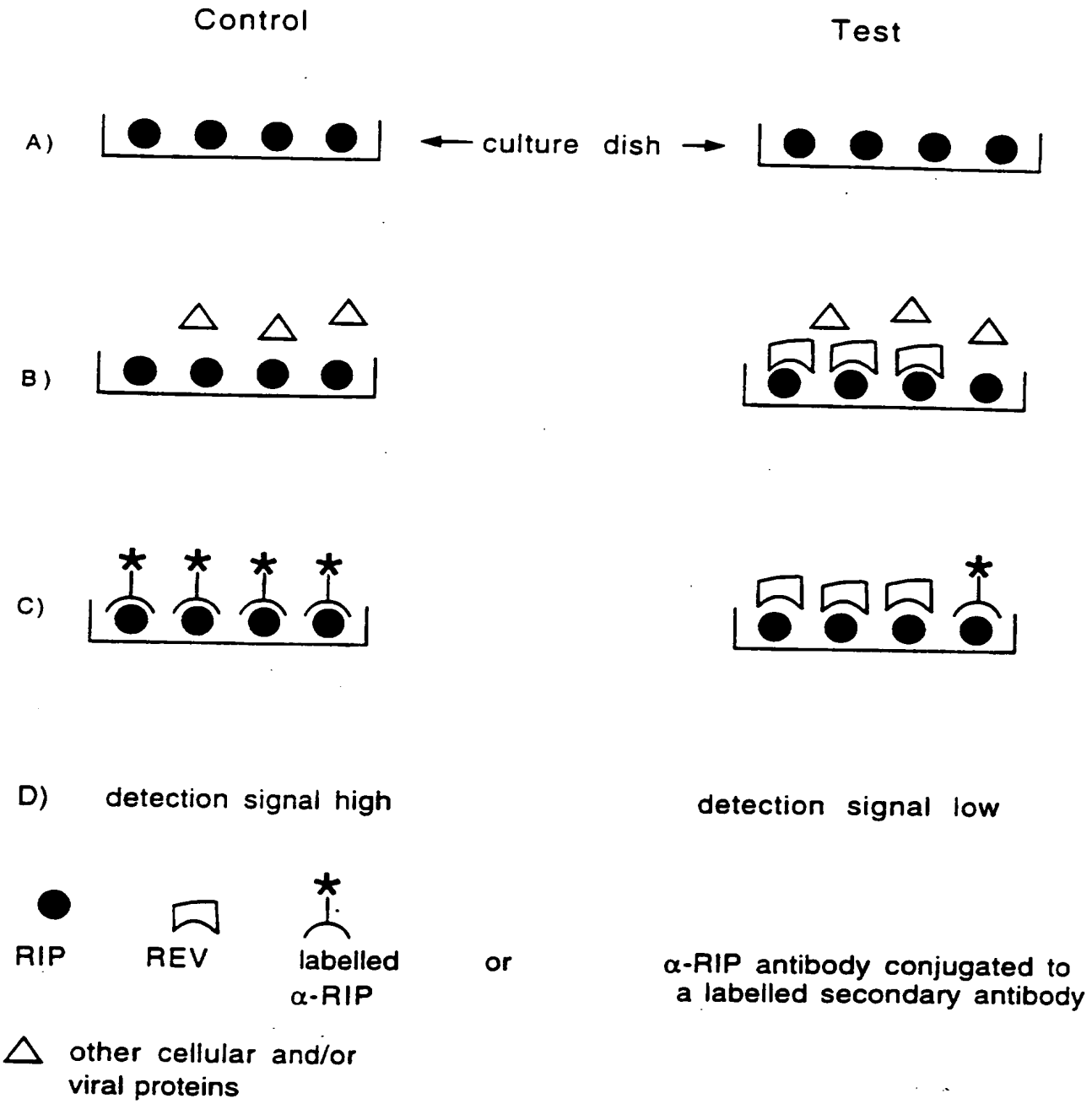


FIG. 8

# INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/09589

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(6) :C12Q 1/02, 1/68
US CL : 435/6, 7.31, 7.8, 29
According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 7.31, 7.8, 29

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, Dialog
search terms: two-hybrid, peptide, scaffold, fibronectin, cyclophilin, zinc finger, G protein

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y, P | Nucleic Acids Research, Vol. 23, No. 7, issued 11 April 1995, Yang et al., "Protein-peptide interactions analyzed with the yeast two-hybrid system", pages 1152-1156, see the entire document. | 1-5 |
| Y | Trends in Genetics, Vol. 10, No. 8, issued August 1994, Fields et al., "The two-hybrid system: an assay for protein-protein interactions", pages 286-292, see page 290. | 1-5 |
| Y | US, A, 5,283,173 (FIELDS ET AL.) 01 February 1994, see columns 3 and 4. | 1-5 |
| Y, P | US, A, 5,432,018 (DOWER ET AL.) 11 July 1995, see columns 4 and 8. | 1-5 |

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be part of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 17 OCTOBER 1995 | 01 NOV 1995 |
| Name and mailing address of the ISA/US<br>Commissioner of Patents and Trademarks<br>Box PCT<br>Washington, D.C. 20231<br>Facsimile No. (703) 305-3230 | Authorized officer<br>ERIC GRIMES<br>Telephone No. (703) 308-0196 |

Form PCT/ISA/210 (second sheet)(July 1992)*

# INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/09589

**Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)**

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

**Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-5

**Remark on Protest**

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

# INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/09589

## BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

Group I, claims 1-5, drawn to a method of isolating peptides that bind to a predetermined target protein.
Group II, claims 6 and 11-13, drawn to the DNA of SEQ ID NO:1, a vector comprising said DNA and a method of using said vector.
Group III, claims 7 and 14, drawn to the peptide of SEQ ID NO:2 and a method of using said peptide.
Group IV, claim 8, drawn to the DNA of SEQ ID NO:4.
Group V, claim 9, drawn to the peptide of SEQ ID NO:5.
Group VI, claim 10, drawn to an antibody that binds the peptide of SEQ ID NO:2.
Group VII, claim 15, drawn to a second method of using the peptide of SEQ ID NO:2 (to screen for anti-HIV agents).

The inventions listed as Groups I-VII do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: The claims of the separate Groups do not share corresponding special technical features; i.e. those technical features that define a contribution which each invention, considered as a whole, makes over the prior art. The technical feature of Group I is isolation of target-binding peptides using a two-hybrid system. The technical feature of Group II is the DNA of SEQ ID NO:1. The technical feature of Group III is the peptide of SEQ ID NO:2. The technical feature of Group IV is the DNA of SEQ ID NO:4. The technical feature of Group V is the peptide of SEQ ID NO:5. The technical feature of Group VI is an antibody that binds to the peptide of SEQ ID NO:2. The technical feature of Group VII is a screening assay for anti-HIV agents using the peptide of SEQ ID NO:2.
The two-hybrid system was known in the prior art for identifying protein-binding peptides (see, e.g., Fields et al., Trends in Genetics 10(8):286-292 (1994)). In addition, rev-binding proteins were known in the prior art (see, e.g., Ruhl et al., Journal of Cell Biology 123(6):1309-1320 (1993)). Since the separate Groups of inventions do not share a technical feature that defines a contribution which each invention, considered as a whole, makes over the prior art, the separate inventions do not share a special technical feature. The inventions of the separate Groups therefore lack unity of invention, and a holding of lack of unity for examination purposes is proper.